MCB
Journals.ASM.org

The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers

Deqing Hu, Xin Gao, Marc A. Morgan, Hans-Martin Herz, Edwin R. Smith, Ali Shilatifard

Stowers Institute for Medical Research, Kansas City, Missouri, USA

Histone H3 lysine 4 (H3K4) can be mono-, di-, and trimethylated by members of the COMPASS (complex of proteins associated with Set1) family from *Saccharomyces cerevisiae* to humans, and these modifications can be found at distinct regions of the genome. Monomethylation of histone H3K4 (H3K4me1) is relatively more enriched at metazoan enhancer regions compared to trimethylated histone H3K4 (H3K4me3), which is enriched at transcription start sites in all eukaryotes. Our recent studies of *Drosophila melanogaster* demonstrated that the *Trithorax*-related (Trr) branch of the COMPASS family regulates enhancer activity and is responsible for the implementation of H3K4me1 at these regions. There are six COMPASS family members in mammals, two of which, MLL3 (GeneID 58508) and MLL4 (GeneID 8085), are most closely related to *Drosophila* Trr. Here, we use chromatin immunoprecipitation-sequencing (ChIP-seq) of this class of COMPASS family members in both human HCT116 cells and mouse embryonic stem cells and find that MLL4 is preferentially found at enhancer regions. MLL3 and MLL4 are frequently mutated in cancer, and indeed, the widely used HCT116 cancer cell line contains inactivating mutations in the *MLL3* gene. Using HCT116 cells in which *MLL4* has also been knocked out, we demonstrate that MLL3 and MLL4 are major regulators of H3K4me1 in these cells, with the greatest loss of monomethylation at enhancer regions. Moreover, we find a redundant role between Mll3 (GeneID 231051) and Mll4 (GeneID 381022) in enhancer H3K4 monomethylation in mouse embryonic fibroblast (MEF) cells. These findings suggest that mammalian MLL3 and MLL4 function in the regulation of enhancer activity and that mutations of MLL3 and MLL4 that are found in cancers could exert their properties through malfunction of these Trr/MLL3/MLL4-specific (Trrific) enhancers.

The appearance of diverse cell types with distinct functions during development is largely dependent on cell-specific patterns of gene expression. Enhancers are *cis*-regulatory elements that play an essential role in directing this spatiotemporal gene expression and are well-known for their ability to activate the transcription of their associated genes from a great distance and irrespective of their orientation relative to transcription start sites (1, 2). Genome-wide mapping of histone modifications led to the identification of monomethylation of histone H3 lysine 4 (H3K4me1) as a modification that is enriched at the enhancers (3). Subsequent investigations continued to reveal signatures at enhancers indicative of active as well as poised states that can be distinguished by the presence of histone H3K27 acetylation (H3K27ac) and RNA polymerase II (Pol II) at active enhancers, and histone H3K27 trimethylation (H3K27me3) at poised enhancers, respectively (4–8).

Methylation of H3K4 is an evolutionarily conserved chromatin modification present in all eukaryotes examined thus far. Histone H3K4 can be mono-, di-, or trimethylated. Although H3K4me1 is globally associated with enhancer elements in metazoans, it also flanks the promoters of actively transcribed genes and can be found toward the 3′ end of actively transcribed genes in all eukaryotes (3). Histone H3K4 dimethylation (H3K4me2) can be found at both enhancers and promoters and in the gene bodies of actively transcribed genes (6). Histone H3K4 trimethylation (H3K4me3) is a prominent feature of the promoters of actively transcribed genes but can also be found at poised or low-level-transcribed, bivalently marked promoters in mammalian stem cells in conjunction with H3K27me3 (9–11).

Much of our current knowledge of the implementation of H3K4 methylation comes from the biochemical purification of the *Saccharomyces cerevisiae* Set1 protein within a macromolecular complex named COMPASS (complex of proteins associated with Set1) that is responsible for the implementation of the mono-, di-, and trimethylation of H3K4 (12, 13). In contrast, *Drosophila melanogaster* has three yeast Set1-related H3K4 methylases, dSet1, *Trithorax* (Trx), and *Trithorax*-related (Trr), each of which exists in COMPASS-like complexes (14). Mammals harbor two representatives for each of the three H3K4 methylases found in *Drosophila*, resulting in a total of six COMPASS-like complexes: SET1A and SET1B, related to dSet1; MLL1 (GeneID 4297) and MLL2 (GeneID 9757), related to *Drosophila* Trx; and MLL3 (GeneID 58508) and MLL4 (GeneID 8085), related to *Drosophila* Trr (see Fig. 1A) (15). Although all COMPASS family members share a set of common core subunits, the presence of unique subunits in individual COMPASS-like complexes may confer a regulatory impact and target these complexes to distinct genomic loci to methylate H3K4 in a variety of contexts. Indeed, recent studies have shown that *Drosophila* Set1 and its mammalian homologs, SET1A and SET1B, are responsible for the bulk level of H3K4me2 and H3K4me3 in cells (14, 16–18). Mll1 (GeneID 214162) has been shown to be required for H3K4me3 at the promoters of less than 5% of genes in mouse embryonic fibroblasts (MEFs) (19). Further diversity among the COMPASS-like complexes was revealed by our recent genome-wide analysis demonstrating that

Received 6 September 2013 Returned for modification 15 September 2013
Accepted 25 September 2013
Published ahead of print 30 September 2013
Address correspondence to Ali Shilatifard, ASH@Stowers.org.
Copyright © 2013, American Society for Microbiology. All Rights Reserved.
doi:10.1128/MCB.01181-13

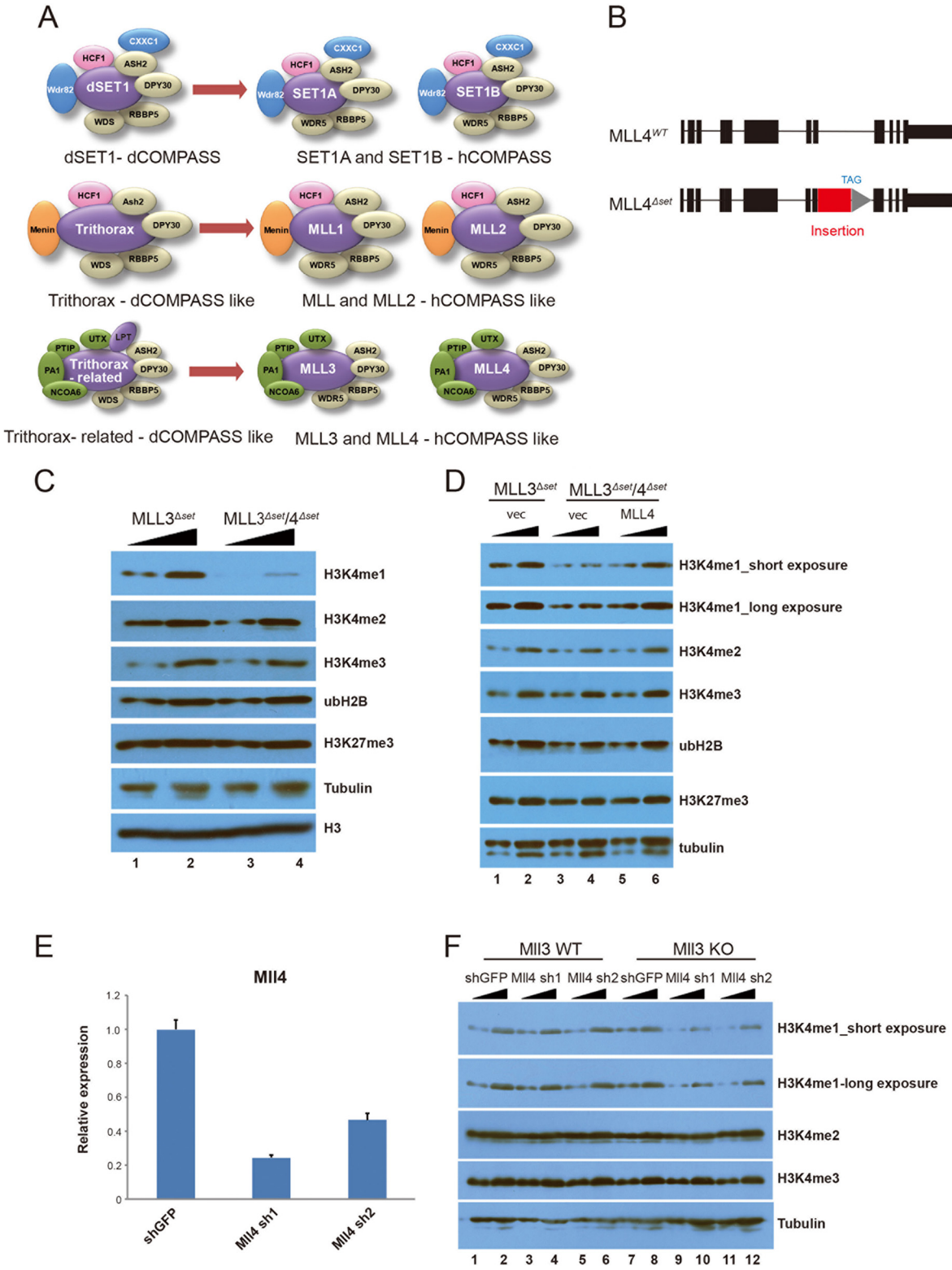

FIG 1 MLL3/MLL4 are required for bulk levels of H3K4me1 in HCT116 cells. (A) The COMPASS family of H3K4 methylases in *Drosophila* and mammals. There are three COMPASS family members in *Drosophila* (dCOMPASS) (14) (left panels). Mammals have six COMPASS members, two members for each *Drosophila* version (middle and right panels) (hCOMPASS, human COMPASS). The subunits common to all COMPASS family members are highlighted in gray. Complex-specific components are represented in blue, orange, and green for the Set1, Trx, and Trr subfamilies, respectively. Hcf1 (pink) is reported to be in Set1 and Trithorax subfamily complexes but not in the Trr family (40). (B) Schematic representation of the *MLL4* alleles in parental HCT116 cells ($MLL4^{WT}$) and *MLL4* disrupted HCT116 cells ($MLL4^{\Delta set}$) as generated previously (22). A DNA fragment harboring a stop codon was inserted upstream of the SET domain-coding exon of the *MLL4* alleles (22). Parental HCT116 cells have a homozygous frameshift mutation before the SET domain of MLL3 (parental HCT116 cells

Mll2 (GeneID 75410) implements H3K4me3 at bivalently marked gene promoters in mouse embryonic stem (ES) cells (20).

Recently, we took advantage of the lower redundancy of the COMPASS members within *Drosophila* to determine that Trr was the enzyme responsible for enhancer-associated H3K4me1 (21). In the present study, we use the colon cancer cell line HCT116 to investigate a potential role in the enhancer function of these enzymes in mammals (22). The cell line HCT116 is homozygously mutated for *MLL3*. To investigate a potential role for MLL4 in enhancer function, we compared chromatin modifications of parental HCT116 cells to published HCT116 cells with a stop codon engineered upstream of the MLL4 SET domain (22, 23). We find that MLL4 is enriched at the enhancer regions and that the loss of MLL4 in a MLL3-deficient background leads to a bulk reduction of H3K4me1. Our genome-wide analysis demonstrates that the loss of MLL3/MLL4 results in the loss of H3K4me1 predominantly at enhancers. In addition, we observed a redundant role between Mll3 and Mll4 in H3K4me1 in MEFs. Since MLL3 and MLL4 are frequently mutated in numerous types of cancer (24–26), we propose that Trrific (Trr/MLL3/MLL4-specific) enhancer malfunction could result in tissue-specific alterations in gene expression underlying cancer pathogenesis.

MATERIALS AND METHODS

Cell culture and antibodies. Parental HCT116 (*MLL3*$^{\Delta set}$) and *MLL4*$^{\Delta set}$ HCT116 cells were previously described (22) and were grown in McCoy's 5A medium with 10% fetal bovine serum (FBS) (Sigma). *Mll3* wild-type mouse embryonic fibroblast (MEF) cells and *Mll3* knockout MEF cells were cultured in Dulbecco modified Eagle medium (DMEM) with 10% FBS. Mouse KH2 embryonic stem cells were grown in feeder-free ESGRO medium (Millipore). Histone H3K4me1, H3K4me2, H3K4me3, and MLL4 polyclonal antibodies were generated in our lab. Ubiquityl-histone H2B (UbH2B) and p300 antibodies are from Cell Signaling Technology (catalog no. 5546) and Santa Cruz (catalog no. SC-585), respectively.

Plasmids and transfection. Human full-length MLL4 cDNA was subcloned into pFN205K EF1a vector with an N-terminal Halo tag between the SgfI and Pme sites (Promega). HCT116 cells were transfected with X-tremeGENE HP (Roche) according to the manufacturer's instructions with the indicated plasmids, and whole-cell extracts were used for Western blotting.

Lentivirus-mediated RNA interference (RNAi). Wild-type MEF cells and *Mll3* knockout MEF cells were infected with lentivirus harboring either green fluorescent protein (GFP) control short hairpin RNA (shRNA) or Mll4 shRNA in the presence of 8 μg/ml Polybrene (Sigma) for 24 h (target sequence for Mll4 sh1, GGAGTTAAAGGCACCTGATGT; target sequence for Mll4 sh2, GAGCCTGGAACTGTAGGAAAT). The infected cells were selected with 2 μg/ml puromycin for an extra 48 h before harvesting for Western blots.

Immunoprecipitation. Parental HCT116 (*MLL3*$^{\Delta set}$) and *MLL4*$^{\Delta set}$ HCT116 cells were washed with phosphate-buffered saline (PBS) twice and lysed in radioimmunoprecipitation assay (RIPA) buffer (20 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, and 1 mM dithiothreitol) containing proteinase inhibitors (Sigma) for 30 min at 4°C. After centrifugation at 13,000 rpm for 30 min, the supernatant was incubated with the indicated antibodies and protein A/G PLUS agarose (Santa Cruz) overnight with gentle rotation at 4°C. The beads were spun down and washed three times with wash buffer (10 mM Tris-HCl [pH 7.4], 1 mM $MgCl_2$, 300 mM NaCl, 10 mM KCl, 0.2% Triton X-100) before boiling in an SDS loading buffer.

ChIP. Chromatin immunoprecipitation (ChIP) samples were prepared as previously described (27). Briefly, cells were cross-linked with 1% formaldehyde for 10 min at room temperature and quenched by the addition of glycine to a final concentration of 0.125 M. The fixed chromatin was fragmented to 200 to 600 bp in length with a Misonix 3000 ultrasonic cell disruptor and used for immunoprecipitation with the indicated antibodies. Chromatin immunoprecipitation-sequencing (ChIP-seq) libraries were prepared with Illumina's TruSeq DNA sample preparation kit.

ChIP-seq analysis. Sequencing data were acquired through the default Illumina pipeline using Casava V1.8. Reads were aligned to the human genome (UCSC hg19) for HCT116 cells and the mouse genome (UCSC mm9) for mouse embryonic stem cells using the Bowtie aligner v0.12.9, allowing unique mapping reads only and allowing up to three mismatches (28). Reads were extended to 150 bases toward the interior of the sequenced fragment and normalized to total reads aligned (counts per million).

Peaks were called for H3K4me1 and H3K27ac using the broad-domain peak detector SICER v1.1 (29) at the false discovery rate (FDR) of $<$1e$-$8, window size of 200, and gap size of 600. Significant change of H3K4me1 enrichment upon MLL4 knockout was identified by Poisson test at a *P* value of $<$1e$-$3. Peak detection of the other samples was done using MACS (model-based analysis of ChIP-Seq) v1.4.2 (30). Associated control samples were used to determine statistical enrichment at a *P* value of $<$1e$-$5 and a FDR of $<$0.05. The enrichment of Mll4 in mouse embryonic stem (ES) cells was determined at a *P* value of $<$1e$-$3. Gene annotations and transcript start site information were from Ensembl 67 with protein-coding, long intergenic noncoding RNA (lincRNA), microRNA (miRNA), snoRNA, and snRNA genes. Peaks at the transcription start site (TSS) region were defined as overlapping the TSS within a 1-kb window. Gene Ontology (GO) term and pathway (KEGG) analyses were performed using DAVID v6.7 (31, 32).

The high-confidence enriched regions were used to depict enrichment profiles. The enrichment profiles are shown as a binary value of enriched/not enriched. Genes were sorted by the distance between the peak summit and the TSS. Gene regions spanning 50 kb on either side of the TSS were binned into 200-bp windows and are shown from 5′ to 3′.

The H3K4me1-enriched regions in parental HCT116 cells were used to depict coverage profiles. The coverage profiles were separated into TSS and putative enhancer groups by H3K4me1 peak locations and sorted by the H3K4me1 occupancy. The profiles spanning 5 kb on either side of the TSS or H3K4me1 peak center were binned into 25-bp windows.

GEO accession numbers. For mouse ES cells, H3K4me1, H3K27ac, and p300 ChIP-seq data are from GEO accession number GSE24164 (5). H3K4me3 ChIP-seq data are from GEO accession number GSE12241 (33). For human HCT116 cells, H3K27ac ChIP-seq data are from GEO accession number GSE38447 (34). Data in the present study are deposited at GEO under GEO accession number GSE51176.

were denoted as *MLL3*$^{\Delta set}$ in subsequent figures), so the *MLL4*$^{\Delta set}$ HCT116 cells lack methyltransferase activity for both MLL3 and MLL4 (designated *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ in subsequent figures). (C) Deficiency of MLL4 results in a large reduction of H3K4me1 levels in HCT116 cells. Whole-cell lysates isolated from parental HCT116 cells and *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells were analyzed by Western blotting using the indicated antibodies. *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells show bulk reduction of H3K4me1 levels without a detectable effect on other histone modification levels globally. (D) Ectopic expression of MLL4 in *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells restores the H3K4me1 levels similar to those in the parental HCT116 cells. vec, vector. (E) Mll4 mRNA levels after lentivirus-mediated knockdown in mouse embryonic fibroblasts (MEFs). Mll4 expression was determined with quantitative reverse transcription-PCR and is normalized to β-actin (*Actb*). (F) Mll3 functions redundantly with Mll4 in H3K4 monomethylation in MEFs. Mll4 was knocked down in wild-type and *Mll3* knockout (KO) MEFs, respectively, and whole-cell lysates were analyzed by Western blotting with the indicated antibodies. Note that Mll4 knockdown reduces H3K4me1 level in *Mll3* knockout MEFs, while its reduction in WT cells demonstrates no detectable effect on H3K4me1 levels.

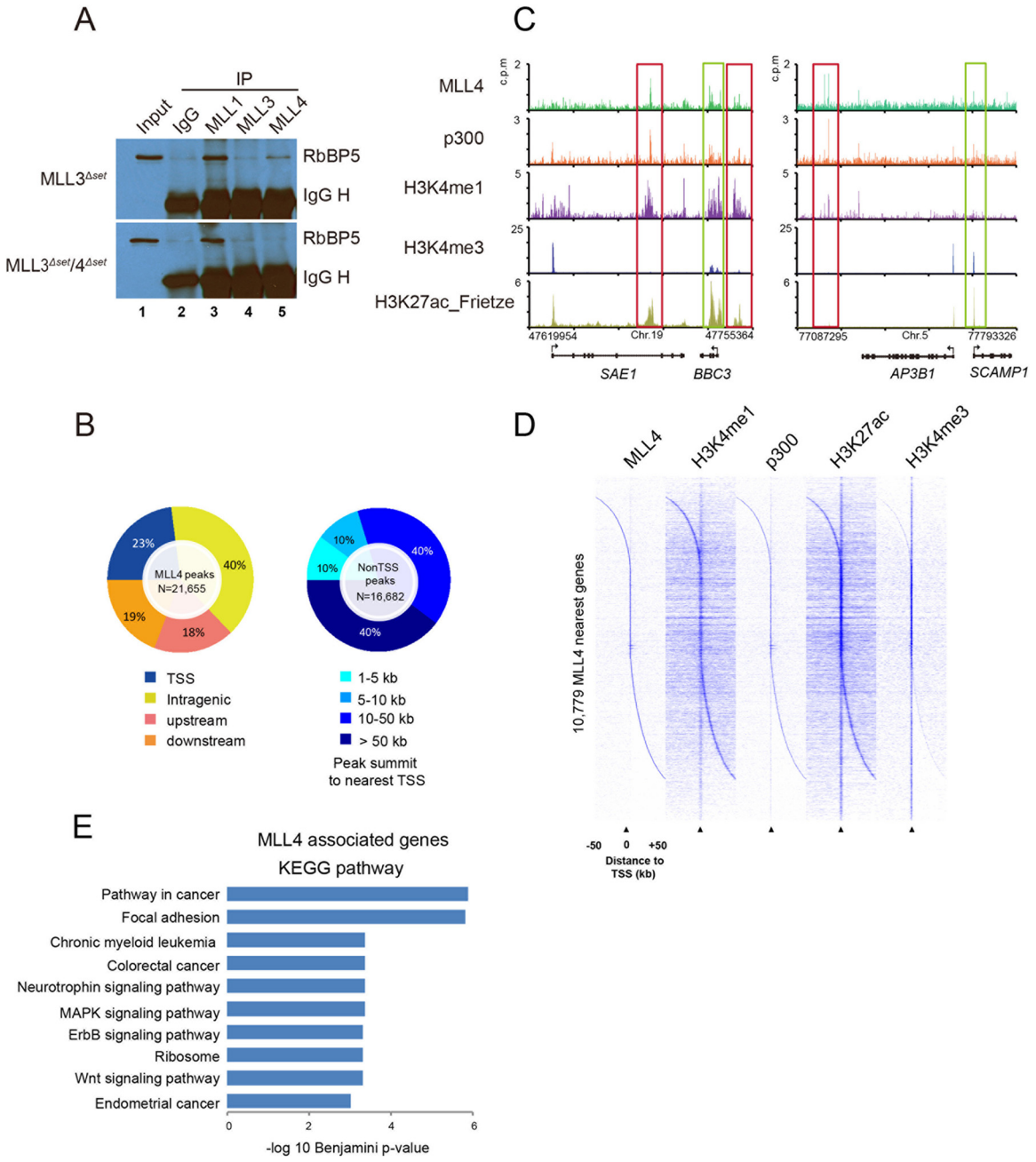

FIG 2 MLL4 is preferentially associated with enhancer regions in HCT116 cells. (A) An MLL4 antibody was tested in immunoprecipitation (IP) to study the disruption pattern for MLL4. Note that the MLL4 antibody is capable of immunoprecipitating the core COMPASS subunit RBBP5 in the parental HCT116 cells but not in the *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells. For a control, we observed that antibodies to MLL (MLL1) are capable of immunoprecipitating RBBP5 in both the parental HCT116 cells and the *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells. The *MLL3* gene is disrupted in HCT116 cells, and accordingly, antibodies to MLL3 can only bring down the background levels of RBBP5. (B) Since our studies in panel A confirmed the specificity of MLL4 antibodies, we tested MLL4 distribution pattern in HCT116 cells using these antibodies. (Left) Pie chart showing the percentage of MLL4 peaks that overlap the transcription start site (TSS), within a gene (Intragenic), or upstream or downstream of the nearest gene. (Right) Positions of MLL4 non-TSS peak summit to the TSS of nearest genes. (C) A representative genomic snapshot of MLL4 peaks showing that MLL4 colocalizes with p300 and H3K4me1 at enhancer regions (red box) or overlaps with H3K4me3 in promoter regions (green box). Histone H3K27ac ChIP-seq data in HCT116 cells are from Frietze et al. (34). (D) Enrichment of binding profiles for MLL4, p300, H3K4me1, H3K4me3, and H3K27ac are shown within 50 kb around the TSS of the MLL4 nearest gene. (E) KEGG pathway analysis of the nearest genes to MLL4 peaks.

RESULTS

MLL3 and MLL4 are responsible for the implementation of bulk levels of H3K4me1. To gain insight into the contribution of the MLL3/MLL4 branch of the COMPASS family to H3K4 methylation in mammals, we chose HCT116 cells since previous studies demonstrated that the *MLL3* gene was inactivated by a frameshift mutation in this cell line upstream of the ePHD, FYR, and SET domains (23). Subsequently, both copies of *MLL4* were inacti-

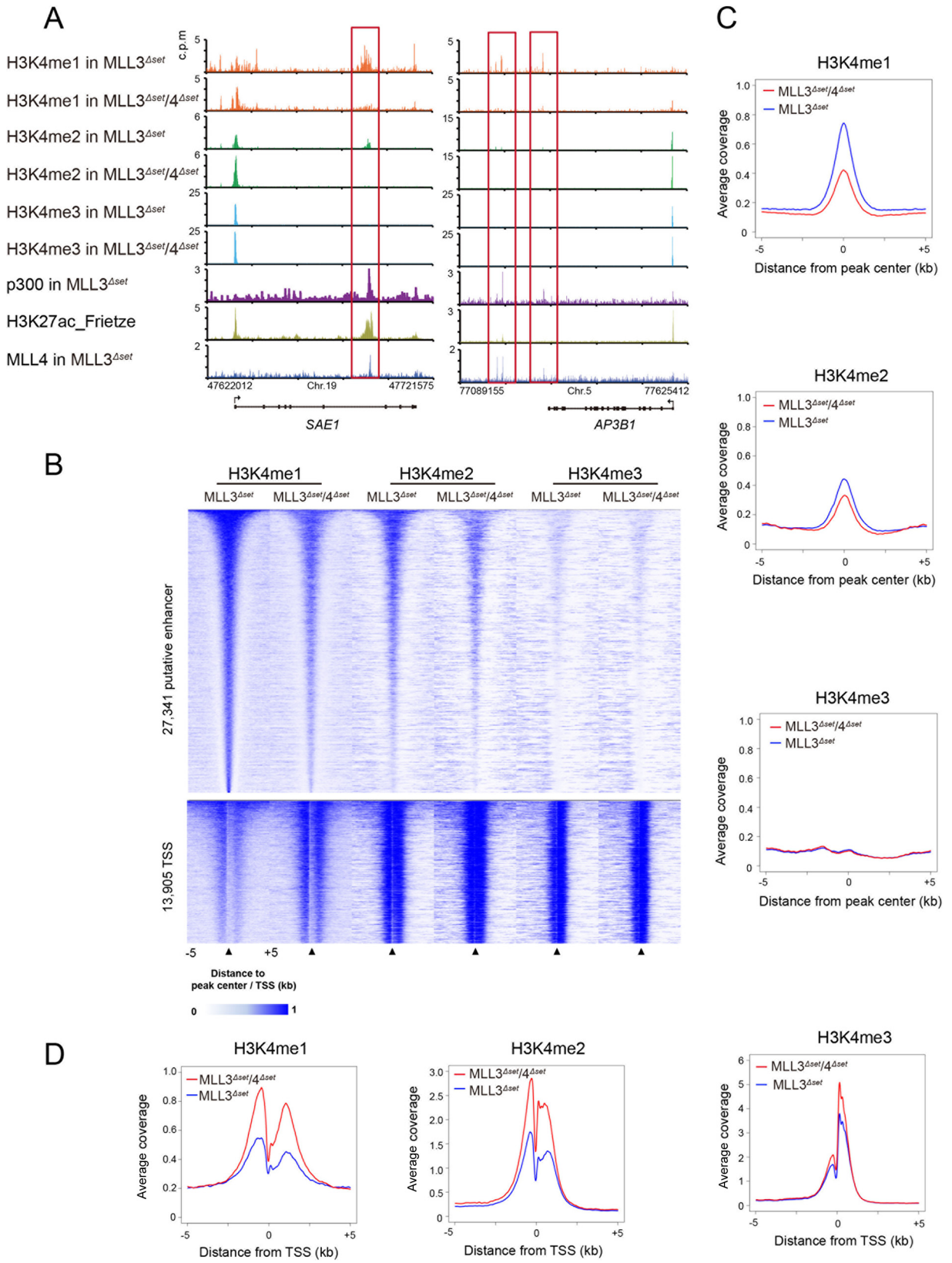

FIG 3 Histone H3K4me1 is diminished at enhancer regions in the absence of MLL4. (A) Genome browser track examples for H3K4me1, H3K4me2, and H3K4me3 occupancy in the parental HCT116 and *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells. Published ChIP-seq data for H3K27ac in HCT116 cells (34) are used to help identify enhancer regions. Chr., chromosome. (B) Coverage profiles of H3K4me1, H3K4me2, and H3K4me3 in parental HCT116 and *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells. A total of 27,341 putative enhancers and 13,095 TSS with H3K4me1 enrichment in parental HCT116 cells are shown. A region within 5 kb around the center of each putative enhancer (top panel) or transcription start site (TSS) (bottom panel) is displayed. Regions are sorted from highest to lowest H3K4me1 occupancy in parental HCT116 cells. (C) Quantitation of the H3K4 methylation states for the same 5-kb window around putative enhancers from panel B. (D) Quantitation of the H3K4 methylation states for the same 5-kb window around the TSS from panel B.

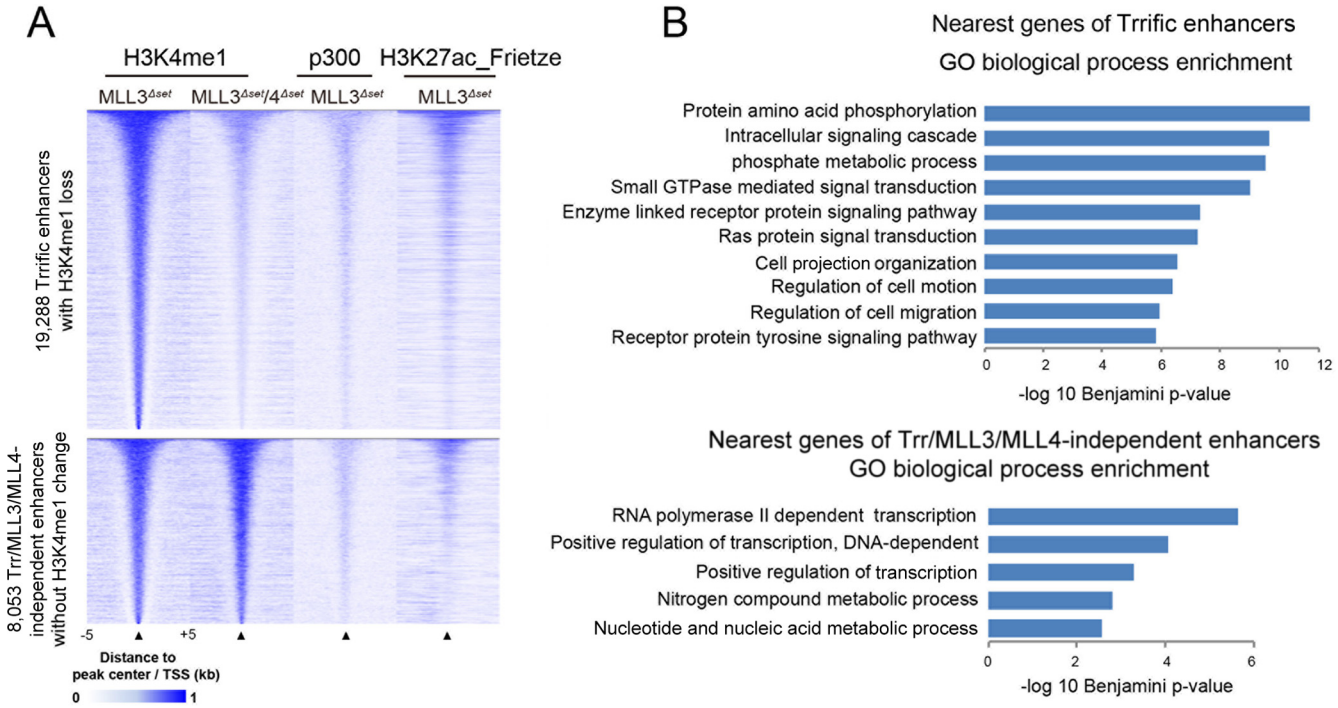

FIG 4 Existence of MLL3/MLL4-independent enhancers in HCT116 cells. (A) Coverage profiles showing a large percentage of putative enhancers exhibit a significant loss in H3K4me1 ($P < 1e-3$) in the absence of MLL3/MLL4 (we call these enhancers Trr/MLL3/MLL4-specific [Trrific] enhancers), while a minority of putative enhancers is independent of MLL3/MLL4 (we call these enhancers Trr/MLL3/MLL4-independent enhancers). (B) Gene ontology (GO) analysis of genes closest to Trrific enhancers and Trr/MLL3/MLL4-independent enhancers. Benjamini-corrected *P* values are shown.

vated in this cell line by inserting a stop codon-containing cassette upstream of the SET domain-coding sequence (22) (Fig. 1B). Therefore, parental HCT116 cells can be designated $MLL3^{\Delta set}$ cells and the cells with the stop codon inserted in the *MLL4* gene can be designated $MLL3^{\Delta set}/4^{\Delta set}$ cells.

We examined levels of mono-, di-, and trimethylated H3K4 and other histone modifications in $MLL3^{\Delta set}$ and $MLL3^{\Delta set}/4^{\Delta set}$ cells using antibodies specifically recognizing different H3K4 methylation states. We find a substantial reduction in the bulk level of H3K4me1 in the absence of active MLL4, while the bulk level of other tested histone modifications are not greatly affected (Fig. 1C, lanes 1 and 2 versus lanes 3 and 4). To further confirm that the absence of MLL4 was primarily responsible for the observed decrease of H3K4me1, we reintroduced MLL4 into $MLL3^{\Delta set}/4^{\Delta set}$ cells. Ectopic expression of MLL4 restores the H3K4me1 level to a level comparable to that in $Mll3^{\Delta set}$ cells (Fig. 1D, lanes 1 to 4 versus lanes 5 and 6). To test whether MLL3 also plays a similar role in histone H3K4me1, we depleted Mll4 levels in wild-type MEF and *Mll3* knockout MEF cells, respectively (Fig. 1E). Our results demonstrated that the *Mll3* knockout MEF has H3K4me1 levels identical to those of wild-type MEFs (Fig. 1E, lanes 1 and 2 versus lanes 7 and 8). However, knockdown of Mll4 in wild-type MEFs demonstrated no detectable effects on the bulk level of H3K4me1, while Mll4 depletion in *Mll3* null MEFs resulted in a decrease in global levels of H3K4me1 (Fig. 1F, lanes 1 and 2 versus lanes 3 to 6 and lanes 7 and 8 versus lanes 9 to 12). These results suggest that Mll3 has some redundancy with Mll4 in monomethylating H3K4 in these cells. In summary, these results demonstrate that MLL3/MLL4 function similarly to *Drosophila* Trr as a major monomethylase of H3K4.

MLL4 is preferentially enriched at enhancer regions in HCT116 cells lacking MLL3. To investigate MLL4's occupancy throughout the genome, we first identified MLL4 binding sites in a genome-wide manner by performing ChIP-seq using antibodies we have developed for MLL4. We demonstrated that our MLL4 antibodies can immunoprecipitate the core COMPASS subunit RBBP5 in $MLL3^{\Delta set}$ cells; however, in $MLL3^{\Delta set}/4^{\Delta set}$ cells, only background levels of RBBP5 were detected (Fig. 2A, lane 5). As the *MLL3* gene in the HCT116 cells was inactivated by a homozygous frameshift mutation before the SET domain (23), MLL3 antibodies fail to coimmunoprecipitate RBBP5 from $MLL3^{\Delta set}$ or $MLL3^{\Delta set}/4^{\Delta set}$ cell extracts (Fig. 2A, lane 4). In contrast, MLL1 antibodies could coimmunoprecipitate RBBP5 both in $MLL3^{\Delta set}$ cells and $MLL3^{\Delta set}/4^{\Delta set}$ cells (Fig. 2A, lane 3).

Our genome-wide analyses identified 21,655 peaks for MLL4 in HCT116 cells (Fig. 2B). Analysis of the peak distribution shows that 77% of the MLL4 peaks are enriched at intergenic and intragenic regions, while only 23% of the peaks overlap transcription start sites (TSS) (Fig. 2B, left panel). Moreover, the majority of intergenic and intragenic MLL4 peaks are located at large distances from an annotated TSS (Fig. 2B, right panel). Further inspection of individual intergenic and intragenic MLL4 peaks at *SAE1* and *AP3B1* loci reveals that MLL4 cooccupies sites together with H3K4me1, p300, and H3K27 acetylation, all of which are used as markers of enhancer regions (3, 5, 35) (Fig. 2C). Binding profiles of the 21,655 MLL4-enriched regions to the nearest gene in HCT116 cells show that MLL4 generally cooccurs with H3K4me1, p300, and H3K27ac, and away from the TSS, and only a small percentage of the MLL4 peaks overlap the highly enriched

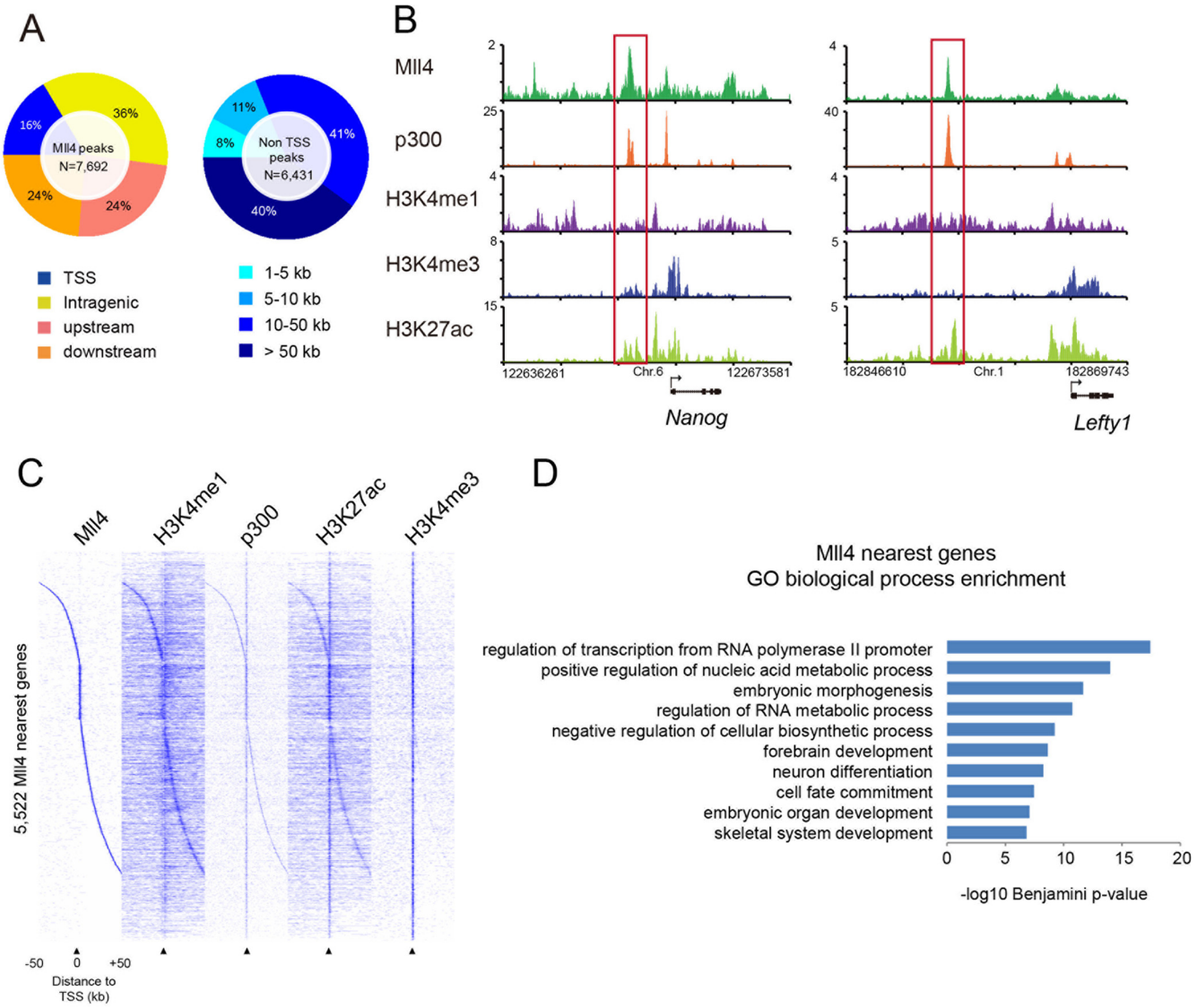

FIG 5 Mll4 occupies enhancers in mouse embryonic stem cells. (A) Pie charts showing the genomic distribution of Mll4 peaks. (Left) Percentages of Mll4 peaks that overlap transcription start sites (TSS), reside within gene bodies, or are upstream or downstream of the nearest genes are indicated. (Right) Positions of Mll4 non-TSS peak summit to the TSS of the nearest genes. (B) Genome browser track examples showing Mll4 binding to regions surrounding *Nanog* and *Lefty1* loci in mouse embryonic stem cells. Previously described enhancer regions are shown outlined in a red box. Numbers on the *y* axis are in counts per million. (C) Enrichment profiles showing a global view of Mll4 binding sites to their 5,522 nearest genes. Since our antibodies were generated toward human MLL4, its strength in mouse cells is lower than its strength in human cells. The high-confidence peaks for Mll4, H3K4me1, p300, H3K27ac, and H3K4me3 are shown for regions 50 kb upstream and downstream of the TSS for the genes nearest Mll4 in mouse ES cells. (D) GO term enrichment analysis for the genes nearest Mll4 peaks identifies several biological processes associated with development.

H3K4me3 found at transcription start sites (Fig. 2D). KEGG pathway enrichment analysis for genes nearest the MLL4 peaks shows enrichment for pathways in cancer (Fig. 2E).

MLL4 is required for the implementation of H3K4me1 at enhancers in human cells. To examine a role for MLL4 at enhancers, we performed ChIP-seq of H3K4me1, H3K4me2, and H3K4me3 in *MLL3*$^{\Delta set}$ and *MLL3*$^{\Delta set}$/*4*$^{\Delta set}$ cells (Fig. 3A to D). We find that H3K4me1 levels are reduced as the result of MLL3/MLL4 deletion at many intergenic and intragenic regions resembling enhancers as exemplified by the genomic regions surrounding the *SAE1* and *AP3B1* loci (Fig. 3A). Since the ratio of H3K4me1 to H3K4me3 can distinguish enhancers from TSS (36), we identified 27,341 H3K4me1 peaks at the putative enhancers (Fig. 3B, top panel) compared to 13,905 peaks at TSS (Fig. 3B, bottom panel). Histone H3K4me1 and, to a lesser extent, H3K4me2 are reduced at enhancer regions in the absence of MLL4 (Fig. 3B and C). Interestingly, H3K4me1 and H3K4me2 show an increase at regions surrounding TSS in the absence of MLL4 (Fig. 3B and D). This is particularly noteworthy, as we saw a similar phenomenon in *Drosophila* S2 cells, where loss of Trr led to increased H3K4me1 at TSS (21). Although we still lack any explanation for this phenomenon, it is conceivable that MLL4 and Trr could restrict the methyltransferase activity of other COMPASS family members at TSS, perhaps via enhancer-promoter communications or by directly bind-

ing to TSS regions, and thereby, preventing other COMPASS-like complexes from methylating those nucleosomes. However, other, more-indirect effects are also possible.

A minority of putative enhancer H3K4 monomethylation is independent of MLL3/MLL4. To quantitatively support a major role for MLL4 in H3K4me1 implementation at enhancers in HCT116 cells, we separated 27,341 putative enhancers into two classes on the basis of significant loss of H3K4me1 ($P < 1 \times 10^{-3}$) (Fig. 4A). Our analysis reveals that 70% of putative enhancers show significant loss in H3K4me1 levels in $MLL3^{\Delta set}/4^{\Delta set}$ cells (19,288 versus 27,341) (Fig.4A, top panel) (we call these enhancers as Trr/MLL3/MLL4-specific [Trrific] enhancers). The remaining 30% of putative enhancers have no loss of H3K4me1 in $MLL3^{\Delta set}/4^{\Delta set}$ cells (Fig. 4A, bottom panel) (we call these enhancers Trr/MLL3/4-independent enhancers), implying that other branches of the COMPASS family may still play a role in enhancer H3K4 monomethylation, perhaps to a lesser extent. Gene ontology (GO) analysis of genes proximal to Trrific enhancers or Trr/MLL3/MLL4-independent enhancers revealed that many of the genes associated with Trrific enhancers are implicated in intracellular signaling, while genes associated with Trr/MLL3/MLL4-independent enhancers demonstrated a tendency toward involvement in gene transcription.

Mll4 is predominantly associated with enhancer regions in mouse embryonic stem cells. In order to rule out the possibility that the preferential binding of MLL4 to enhancers was specific to HCT116 cells, we also performed ChIP-seq for Mll4 in mouse embryonic stem cells. We identified 7,692 high-confidence Mll4 peaks in mouse ES cells, 84% of which are located at intergenic and intragenic regions (Fig. 5A, left panel). Moreover, the majority of intergenic and intragenic Mll4 peaks are more than 10 kb from an annotated TSS (Fig. 5A, right panel), similar to what we observed of MLL4 binding in HCT116 cells (Fig. 2). Inspection of well-characterized enhancers, such as the *Nanog* and *Lefty1* enhancers, shows Mll4 cooccupying enhancers with p300, H3K4me1, and H3K27ac (Fig. 5B). To further explore whether Mll4 is generally localized to the enhancer regions in mouse embryonic stem cells, we aligned the Mll4 peaks with H3K4me1, p300, H3K27ac, and H3K4me3. Binding profiles of these factors indicate that Mll4 cooccupies enhancers with H3K4me1, p300, and H3K27ac in ES cells (Fig. 5C). Gene Ontology analysis of the closest genes to the intergenic and intragenic Mll4 peaks reveals that many of these genes are linked to intracellular signaling cascades, the regulation of transcription and macromolecule biosynthetic processes, tube development, as well as embryonic morphogenesis (Fig. 5D).

DISCUSSION

Histone H3K4me1 was the first histone modification found to have a global association/correlation with enhancer regions (3); additional modifications such as H3K27ac and H3K27me3 were subsequently shown to associate with only a subset of enhancers (4, 5, 7). We identified *Drosophila* COMPASS family member Trr as a major H3K4 monomethylase functioning at the enhancer regions and showed that it is required in enhancer-promoter communication for faithful and tissue-specific gene expression (16, 21). While Trr has long been considered a divergent form of MLL3 and MLL4 with homology restricted to the C-terminal halves of these proteins, it was recently found that a second gene encodes a protein with homology to the N-terminal halves of MLL3 and MLL4 (14, 37). This protein, named Lpt, associates with Trr in a COMPASS-like complex and is also required for bulk H3K4me1 (14, 21). Furthermore, Trr-Lpt COMPASS and MLL3/MLL4 COMPASS share a similar subunit composition (Fig. 1A), suggesting that these complexes should behave similarly to the *Drosophila* Trr-Lpt complex. However, the potential functional redundancy shared among MLL3 and MLL4 had previously precluded a definitive test of this model (21). By using the recently published HCT116 $MLL4^{\Delta set}$ cells and comparing these to the parental HCT116 cells that are already homozygously mutated for *MLL3*, we have been able to demonstrate the following. (i) Enhancer-associated monomethylation of H3K4 in mammals resides mainly in the MLL3/4 branches of the COMPASS family. (ii) MLL3 and MLL4 activity toward H3K4me1 on enhancers appears to be redundant to some extent. (iii) MLL4 is preferentially mapped to the enhancer regions both in HCT116 cells and mouse embryonic stem cells. (iv) MLL4 deficiency results in a global reduction of the H3K4me1 level, which mainly occurs at intergenic and intragenic regions with chromatin signatures of enhancers. Therefore, we conclude that MLL3/4 COMPASS-like complexes possess H3K4 monomethylase activity on enhancers (Fig. 6).

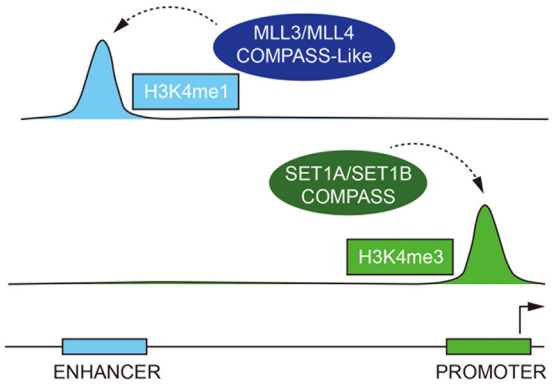

FIG 6 A model for differential genomic localization of histone H3K4me3 and H3K4me1 and implementation by distinct branches of the COMPASS family. The MLL3/4 branches of the COMPASS family are major histone H3K4 monomethylases (H3K4me1) at enhancers. In contrast, SET1A/B-COMPASS branches function as H3K4 trimethylases (H3K4me3) at promoters.

Although H3K4me1 is a well-established enhancer signature, the functional significance of H3K4me1 at these regions is currently unknown. Mutations of MLL3, MLL4, and the associated factor UTX have each been found to frequently occur in multiple types of cancers (24–26, 38). Furthermore, the enhancer chromatin landscape was demonstrated to be altered in colon cancers, and these changes could drive a specific transcriptional program required for colon carcinogenesis (39). Therefore, the identification of the MLL3/4 branch of the COMPASS family as major enhancer monomethylases (Fig. 6) will provide a basis for investigating the normal function of H3K4me1 at enhancers during development and suggests that mutations in MLL3 and MLL4 associated with different forms of cancer may result in enhancer malfunction.

ACKNOWLEDGMENTS

We are grateful to Yiping He (Duke University) for providing published parental and $MLL3^{\Delta set}/4^{\Delta set}$ HCT116 cells.

We thank Rhonda Egidy, Allison Peak, Anoja Perera, and the rest of the Stowers Institute Molecular Biology core for help with Illumina sequencing. We are grateful to the Stowers Institute Tissue Culture core for

assistance with the generation and maintenance of cell lines. We thank Laura Shilatifard and Lisa Kennedy for editorial assistance.

These studies were supported in part by National Cancer Institute grant R01CA150265 to A.S.

REFERENCES

1. **Bulger M, Groudine M.** 2011. Functional and mechanistic diversity of distal transcription enhancers. Cell **144:**327–339.
2. **Ong CT, Corces VG.** 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. **12:**283–293.
3. **Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B.** 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. **39:**311–318.
4. **Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EE.** 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. **44:**148–156.
5. **Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R.** 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. **107:**21931–21936.
6. **Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau JC, Ferrier P, Spicuglia S.** 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. EMBO J. **30:**4198–4210.
7. **Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J.** 2011. A unique chromatin signature uncovers early developmental enhancers in humans. Nature **470:**279–283.
8. **Zentner GE, Tesar PJ, Scacheri PC.** 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. **21:**1273–1283.
9. **Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K.** 2007. High-resolution profiling of histone methylations in the human genome. Cell **129:**823–837.
10. **Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES.** 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell **125:**315–326.
11. **Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA.** 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell **130:**77–88.
12. **Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A.** 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. U. S. A. **98:**12902–12907.
13. **Roguev A, Schaft D, Shevchenko A, Pijnappel WW, Wilm M, Aasland R, Stewart AF.** 2001. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J. **20:**7137–7148.
14. **Mohan M, Herz HM, Smith ER, Zhang Y, Jackson J, Washburn MP, Florens L, Eissenberg JC, Shilatifard A.** 2011. The COMPASS family of H3K4 methylases in Drosophila. Mol. Cell. Biol. **31:**4310–4318.
15. **Shilatifard A.** 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu. Rev. Biochem. **81:**65–95.
16. **Ardehali MB, Mei A, Zobeck KL, Caron M, Lis JT, Kusch T.** 2011. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J. **30:**2817–2828.
17. **Hallson G, Hollebakken RE, Li T, Syrzycka M, Kim I, Cotsworth S, Fitzpatrick KA, Sinclair DA, Honda BM.** 2012. dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. Genetics **190:**91–100.
18. **Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, Shilatifard A.** 2008. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol. Cell. Biol. **28:**7337–7344.
19. **Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, Ge K, Krumlauf R, Shilatifard A.** 2009. Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. Mol. Cell. Biol. **29:**6074–6085.
20. **Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A.** 2013. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. **20:**1093–1097.
21. **Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A.** 2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. **26:**2604–2620.
22. **Guo C, Chang CC, Wortham M, Chen LH, Kernagis DN, Qin X, Cho YW, Chi JT, Grant GA, McLendon RE, Yan H, Ge K, Papadopoulos N, Bigner DD, He Y.** 2012. Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. Proc. Natl. Acad. Sci. U. S. A. **109:**17603–17608.
23. **Watanabe Y, Castoro RJ, Kim HS, North B, Oikawa R, Hiraishi T, Ahmed SS, Chung W, Cho MY, Toyota M, Itoh F, Estecio MR, Shen L, Jelinek J, Issa JP.** 2011. Frequent alteration of MLL3 frameshift mutations in microsatellite deficient colorectal cancer. PLoS One **6:**e23320. doi:10.1371/journal.pone.0023320.
24. **Cleary SP, Jeck WR, Zhao X, Chen K, Selitsky SR, Savich GL, Tan TX, Wu MC, Getz G, Lawrence MS, Parker JS, Li J, Powers S, Kim H, Fischer S, Giundi M, Ghanekar A, Chiang DY.** 31 May 2013. Identification of driver genes in hepatocellular carcinoma by exome sequencing. Hepatology [Epub ahead of print.] doi:10.1002/hep.26540.
25. **Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJ, Connors JM, Hirst M, Gascoyne RD, Marra MA.** 2011. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature **476:**298–303.
26. **Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, Boca SM, Carter H, Samayoa J, Bettegowda C, Gallia GL, Jallo GI, Binder ZA, Nikolsky Y, Hartigan J, Smith DR, Gerhard DS, Fults DW, VandenBerg S, Berger MS, Marie SK, Shinjo SM, Clara C, Phillips PC, Minturn JE, Biegel JA, Judkins AR, Resnick AC, Storm PB, Curran T, He Y, Rasheed BA, Friedman HS, Keir ST, McLendon R, Northcott PA, Taylor MD, Burger PC, Riggins GJ, Karchin R, Parmigiani G, Bigner DD, Yan H, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE.** 2011. The genetic landscape of the childhood cancer medulloblastoma. Science **331:**435–439.
27. **Lee TI, Johnstone SE, Young RA.** 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. **1:**729–748.
28. **Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10:**R25. doi:10.1186/gb-2009-10-3-r25.
29. **Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W.** 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics **25:**1952–1958.
30. **Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS.** 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9:**R137. doi:10.1186/gb-2008-9-9-r137.
31. **Huang DW, Sherman BT, Lempicki RA.** 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. **37:**1–13.
32. **Huang DW, Sherman BT, Lempicki RA.** 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. **4:**44–57.
33. **Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE.** 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature **448:**553–560.

34. **Frietze S, Wang R, Yao L, Tak YG, Ye Z, Gaddis M, Witt H, Farnham PJ, Jin VX.** 2012. Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. Genome Biol. **13:**R52. doi:10.1186/gb-2012-13-9-r52.
35. **Visel A, Rubin EM, Pennacchio LA.** 2009. Genomic views of distant-acting enhancers. Nature **461:**199–205.
36. **Calo E, Wysocka J.** 2013. Modification of enhancer chromatin: what, how, and why? Mol. Cell **49:**825–837.
37. **Chauhan C, Zraly CB, Parilla M, Diaz MO, Dingwall AK.** 2012. Histone recognition and nuclear receptor co-activator functions of Drosophila cara mitad, a homolog of the N-terminal portion of mammalian MLL2 and MLL3. Development **139:**1997–2008.
38. **van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, Edkins S, Hardy C, O'Meara S, Teague J, Butler A, Hinton J, Latimer C, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Cole J, Forbes S, Jia M, Jones D, Kok CY, Leroy C, Lin ML, McBride DJ, Maddison M, Maquire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, Pleasance E, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Collins VP, Ichimura K, Law S, Wong J, Yuen ST, Leung SY, Tonon G, DePinho RA, Tai YT, Anderson KC, Kahnoski RJ, Massie A, Khoo SK, Teh BT, Stratton MR, Futreal PA.** 2009. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat. Genet. **41:**521–523.
39. **Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, Myeroff L, Lutterbaugh J, Jarrar A, Kalady MF, Willis J, Moore JH, Tesar PJ, Laframboise T, Markowitz S, Lupien M, Scacheri PC.** 2012. Epigenomic enhancer profiling defines a signature of colon cancer. Science **336:**736–739.
40. **van Nuland R, Smits AH, Pallaki P, Jansen PW, Vermeulen M, Timmers HT.** 2013. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. Mol. Cell. Biol. **33:**2067–2077.